



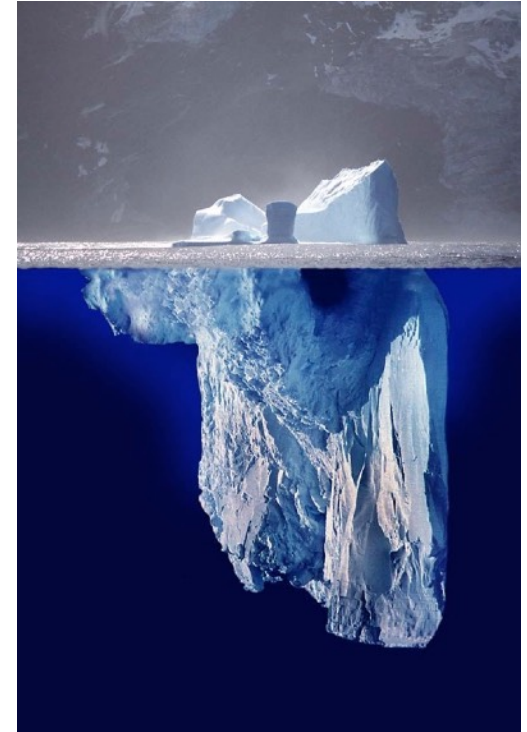
# EBRAINS Provenance API

Andrew Davison  
CNRS / Université Paris-Saclay



# Reproducible research

- ❖ “I thought I used the same parameters but I’m getting different results”
- ❖ “I can’t remember which version of the code I used to generate figure 6”
- ❖ “The new student wants to reuse that model I published three years ago but he can’t reproduce the figures”
- ❖ “It worked yesterday”



# Sharing simulations and data analysis results



The screenshot shows the EBRAINS Knowledge Graph interface. At the top left is the EBRAINS logo. Below it is a search bar with the placeholder text "Search (e.g. brain or neuroscience)". To the right of the search bar are links for "Share data", "About", and "Login". Below the search bar is a "CATEGORIES" section with a table:

Category	Count
Project	122
<b>Dataset</b>	<b>1209</b>
Model	100
Software	150
Contributor	1284

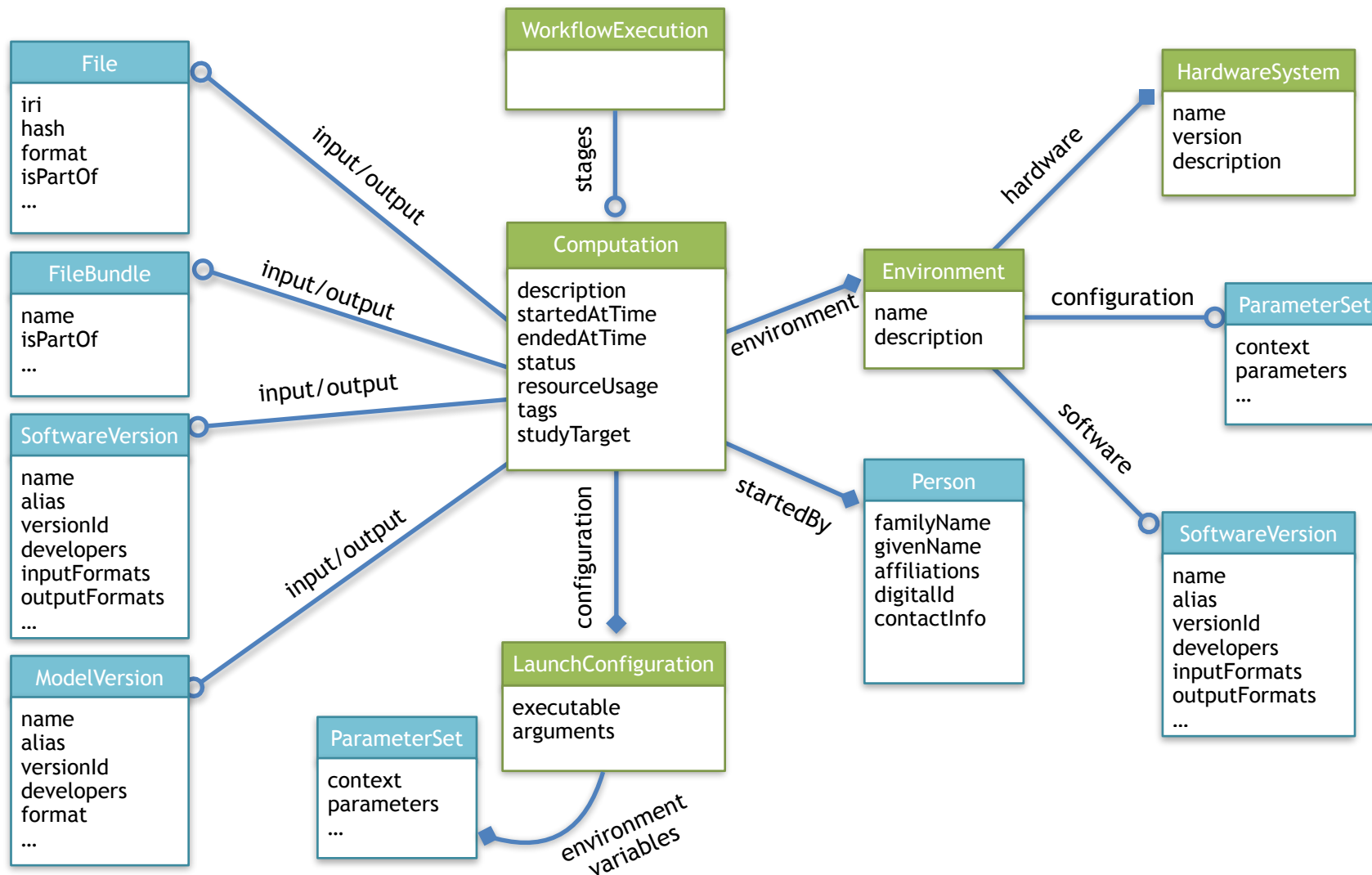
Below the categories is a "FILTERS" section with a "Reset" button. Under "SPECIES", there are checkboxes for "Homo sapiens" (845) and "Mus musculus" (102). The main content area shows "Viewing 1-20 of 1209 results" and a card for "Study of Slow Waves (SWs) propagation through wide-field calcium imaging of the right cortical hemisphere". The card includes a description: "Slow waves (SWs) are spatio-temporal patterns of cortical activity that occur both during natural sleep and anesthesia and are preserved across species. Although electrophysiological recordings have been largely us...", keywords: "Calcium imaging", and methods: "Wide-field fluorescence microscopy". A blue arrow points from the "reproducible results" text in the list to the "Wide-field fluorescence microscopy" method in the card.

- Most datasets in the KG contain “raw” or pre-processed data.
- The KG contains models, but not simulation results.
- We do not currently check that models or code give reproducible results.
- We want to facilitate sharing “derived data” and promote reproducibility.

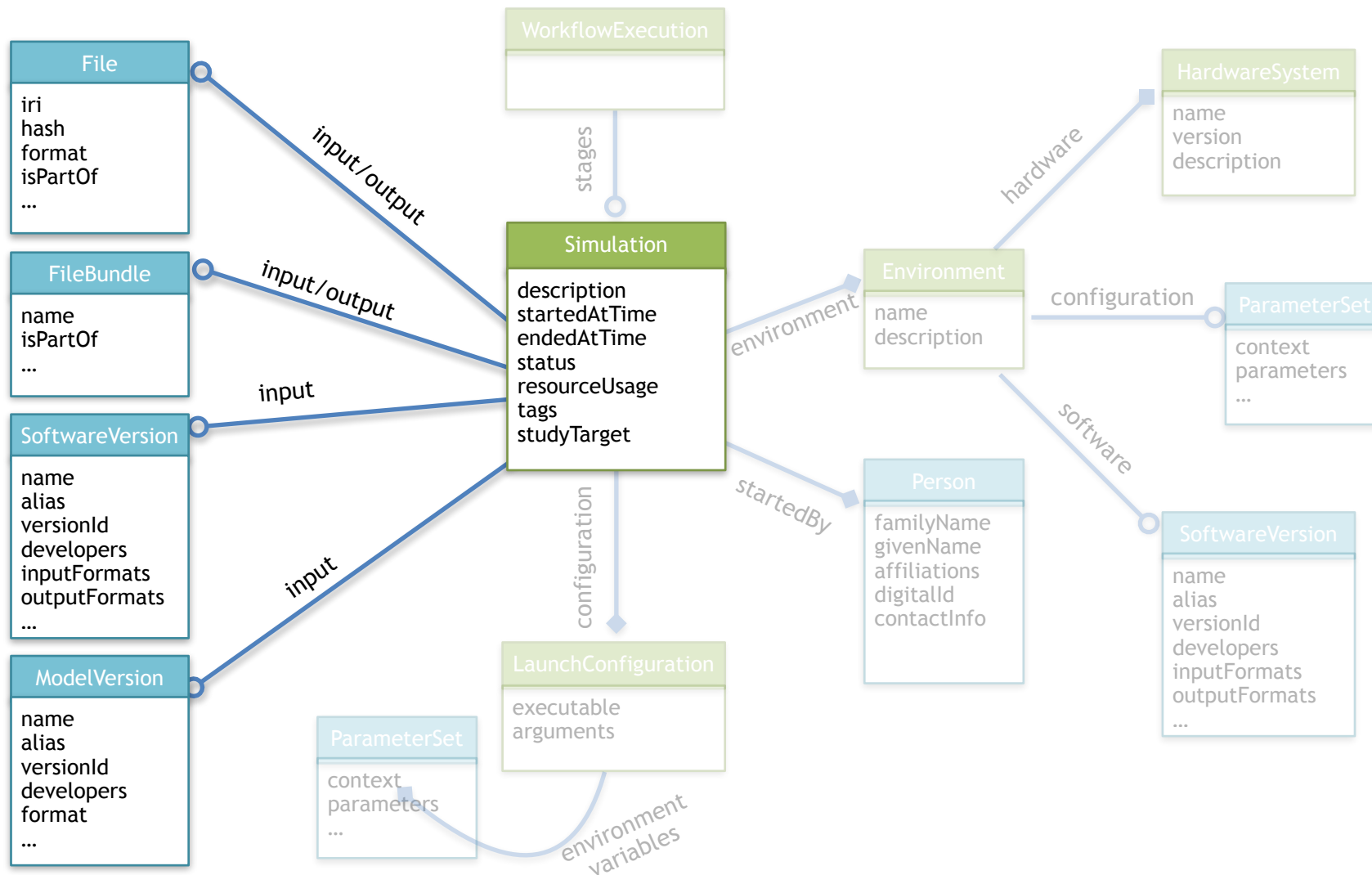
# What information is needed?

- what code was run?
  - which executable?
    - \* name, location, version, compiler, compilation options
  - which script?
    - \* name, location, version
    - \* options, parameters
    - \* dependencies (name, location, version)
- what were the input data?
  - name, location, content
- what were the outputs?
  - data, logs, stdout/stderr
- who launched the computation?
- when was it launched/when did it run? (queueing systems)
- where did it run?
  - machine name(s), other identifiers (e.g. IP addresses)
  - processor architecture
  - available memory
  - operating system
- why was it run?
- what was the outcome?
- which project was it part of?

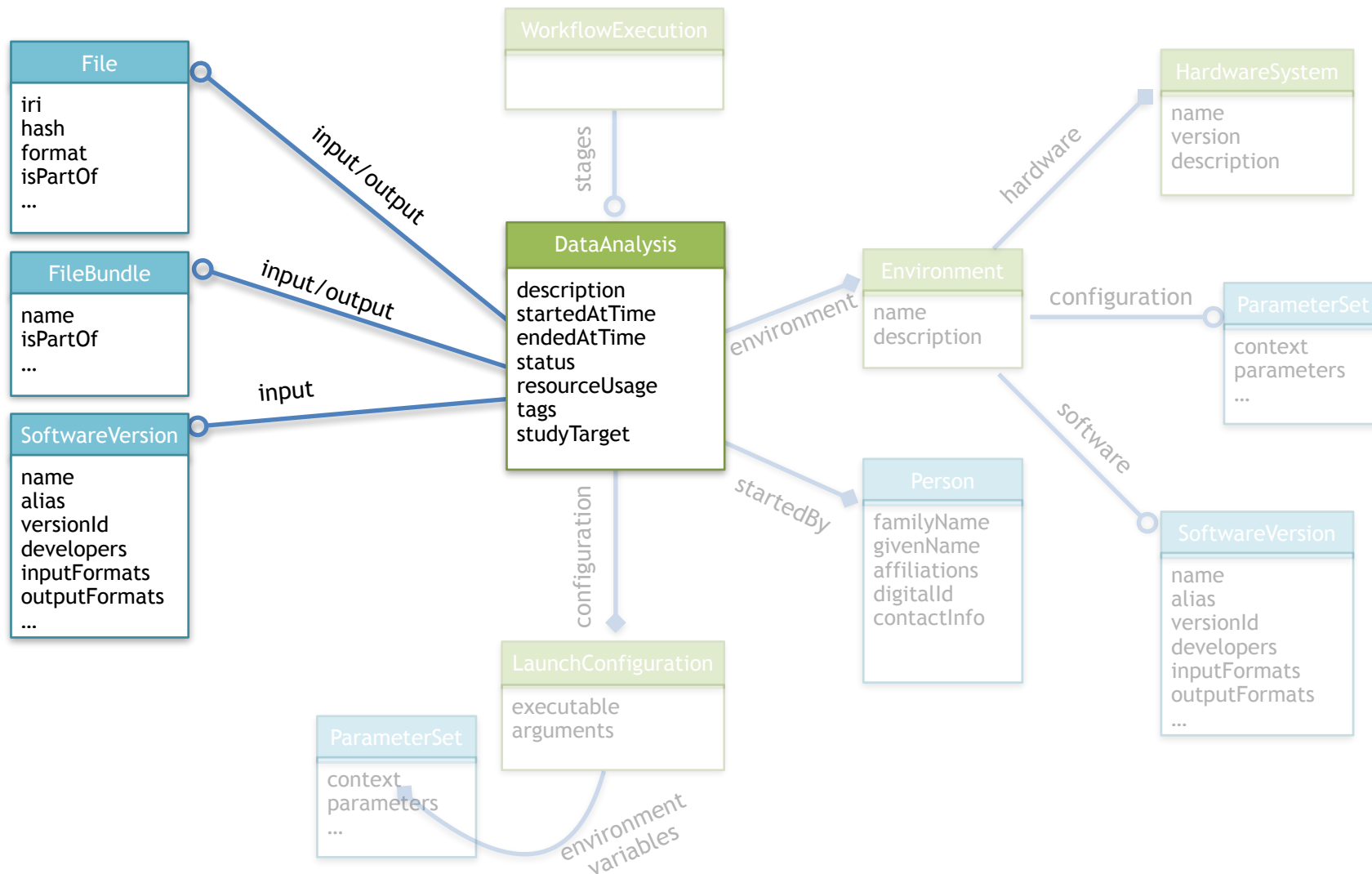
# Knowledge Graph schemas



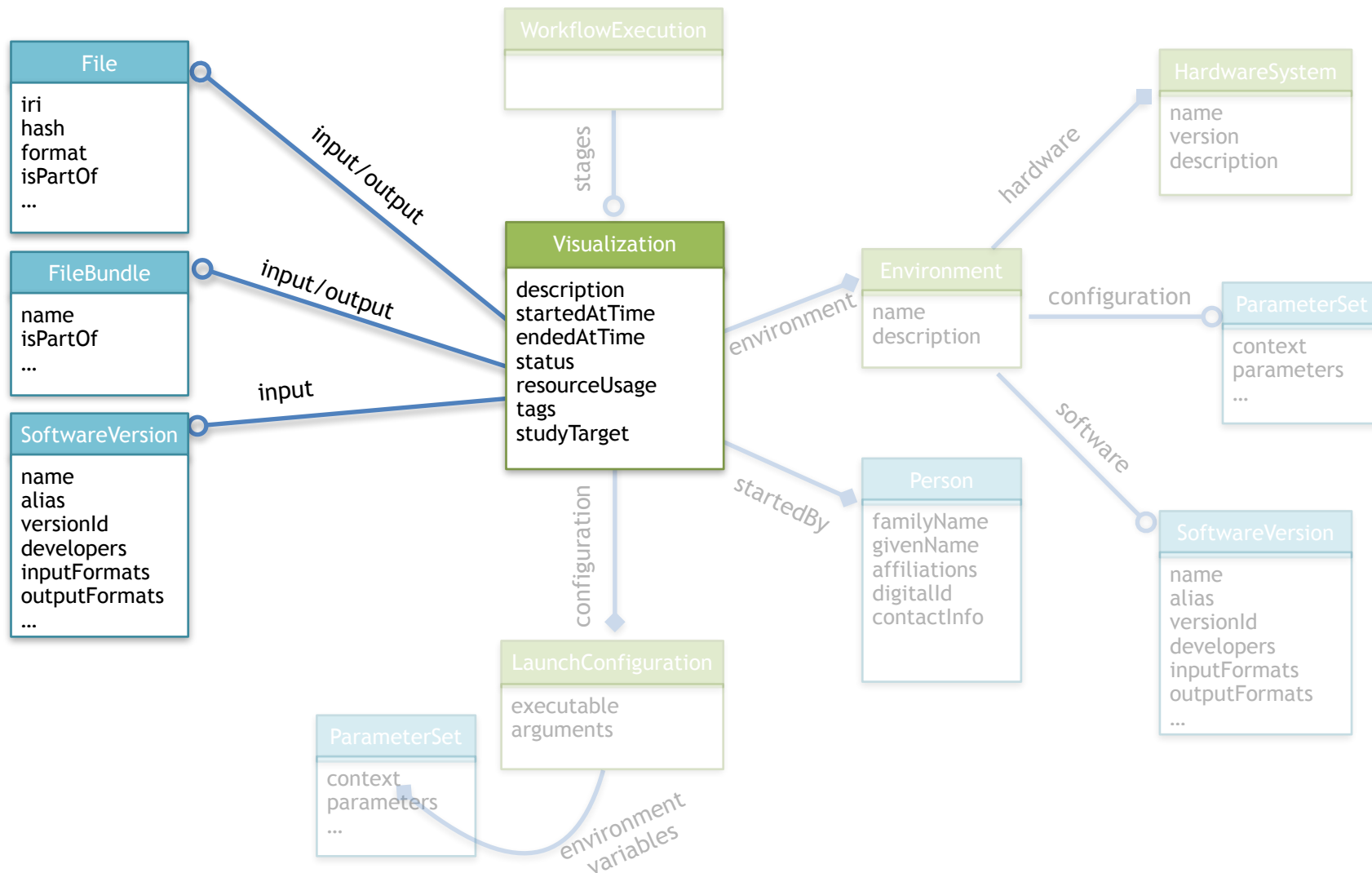
# Knowledge Graph schemas



# Knowledge Graph schemas

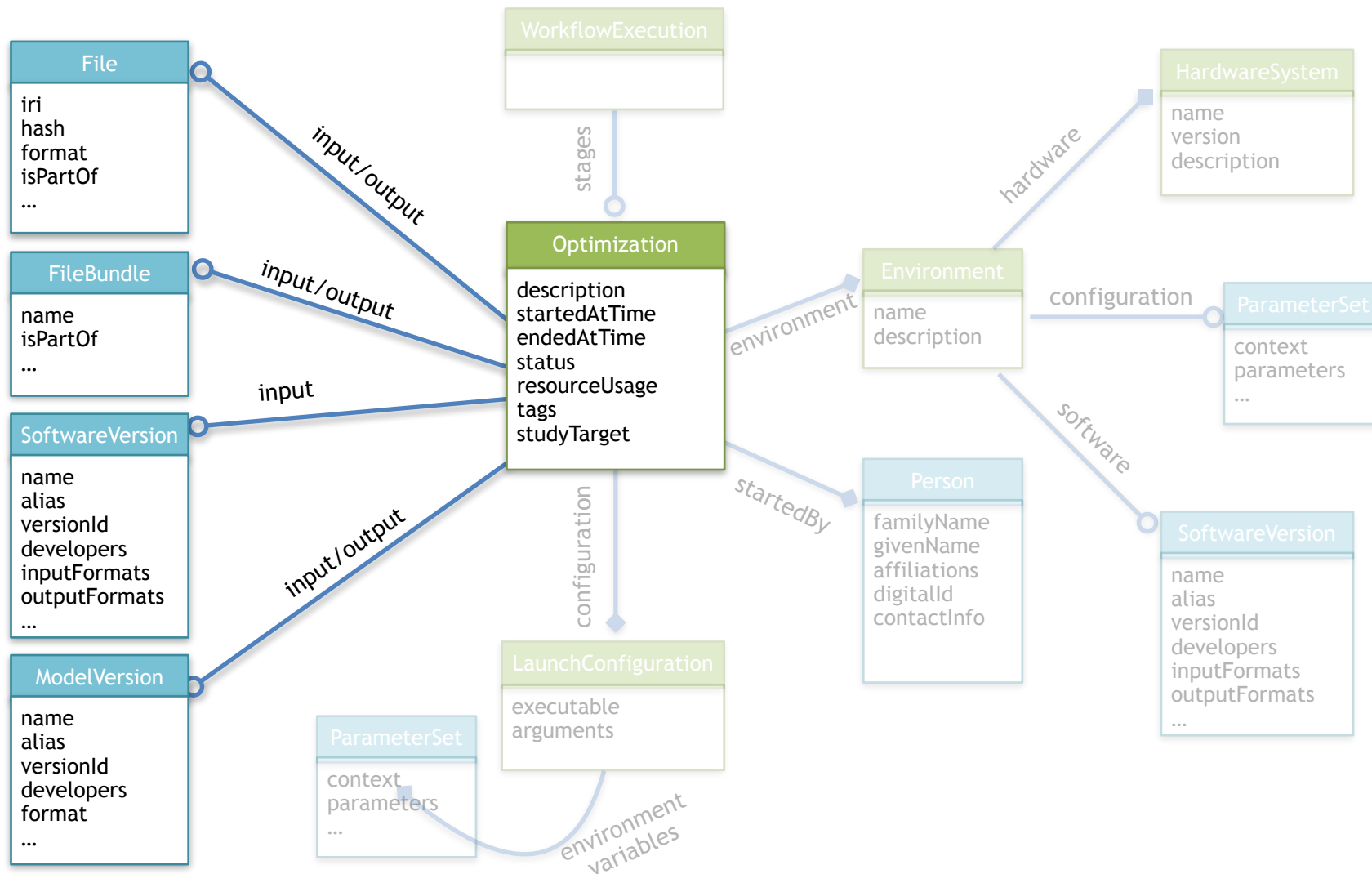


# Knowledge Graph schemas

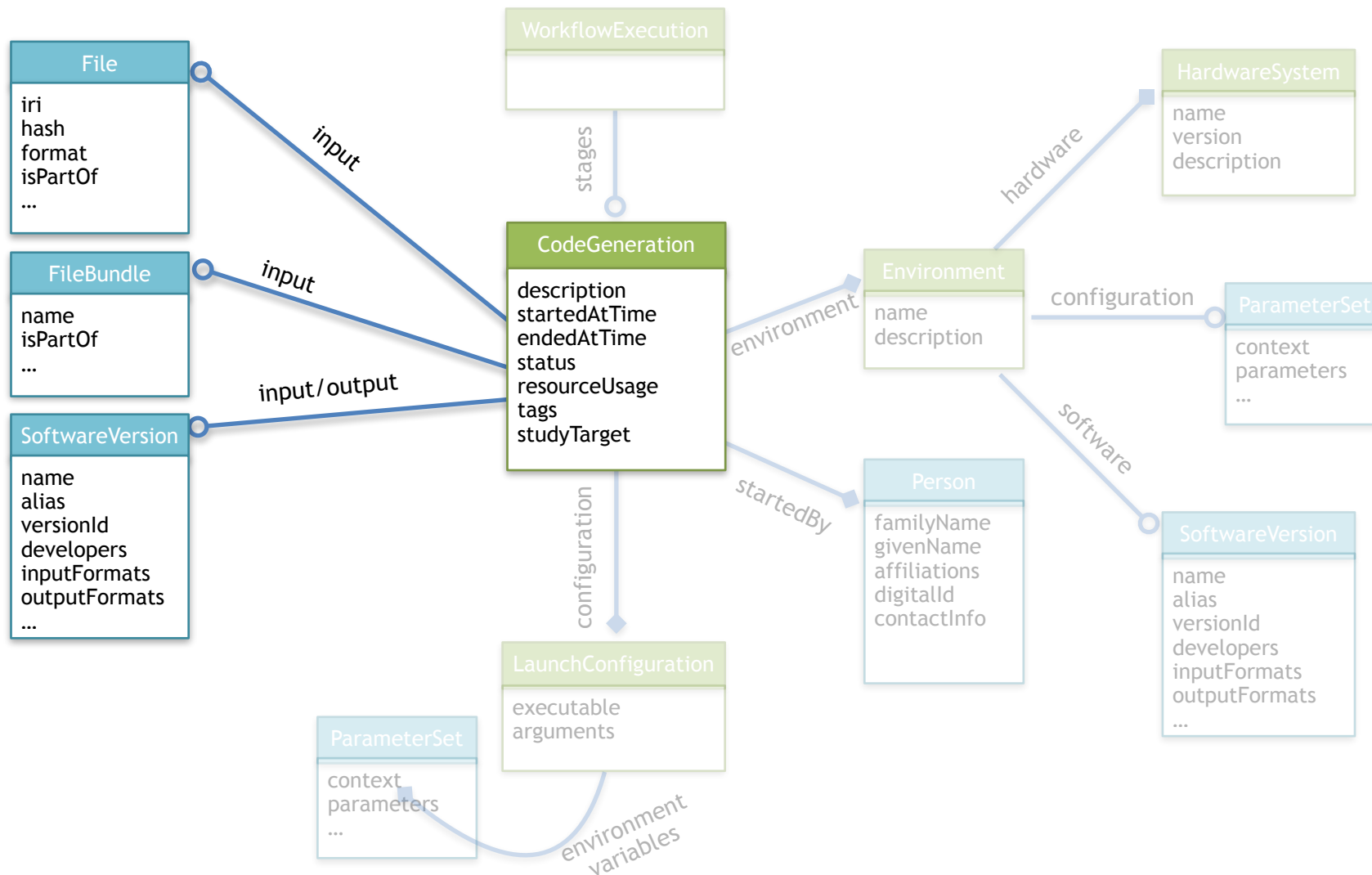




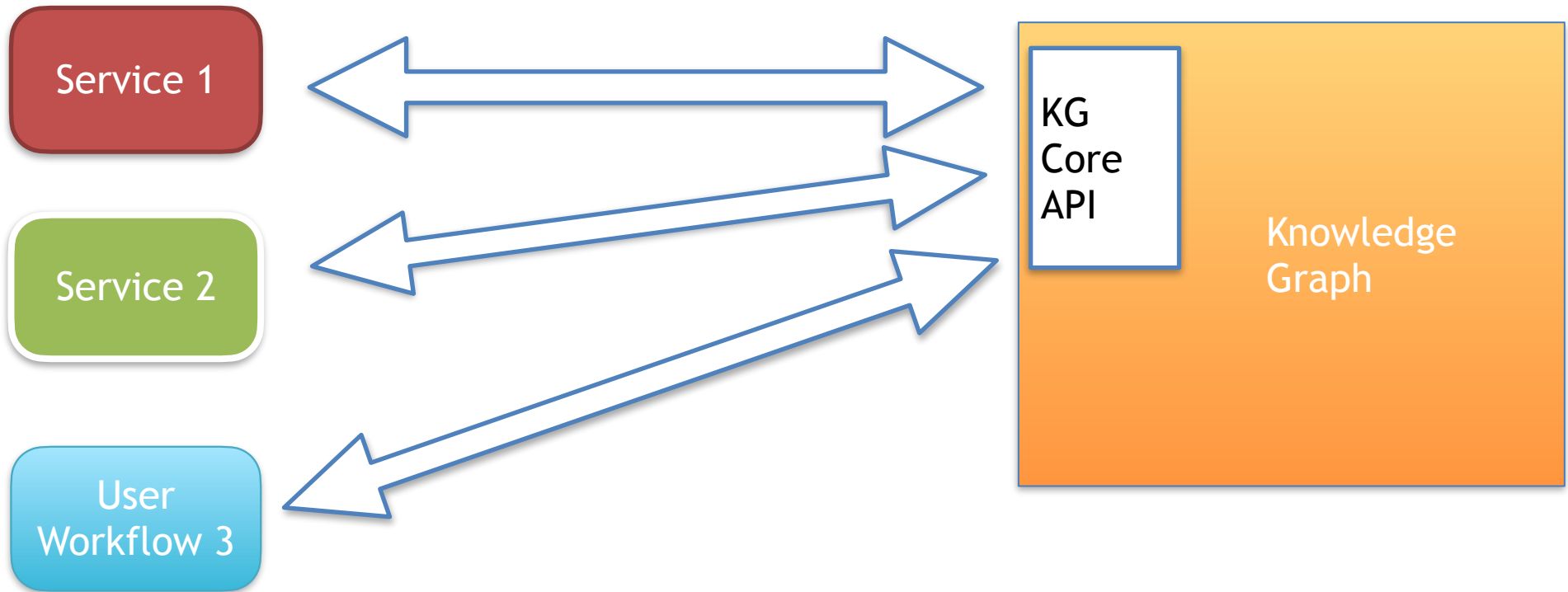
# Knowledge Graph schemas



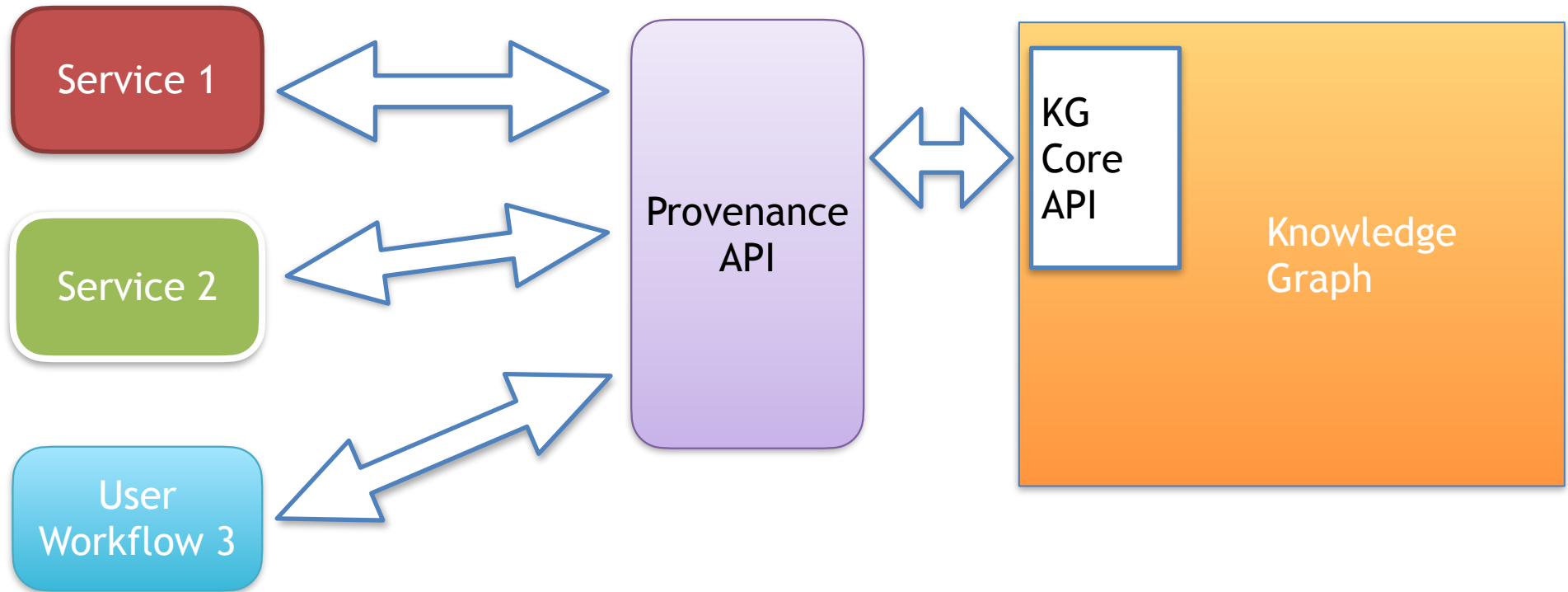
# Knowledge Graph schemas



# Communication between services/tools/workflows and the KG



# Communication between services/tools/workflows and the KG



# Provenance API - α release

## EBRAINS Provenance API <sup>1.0</sup> OAS3

/openapi.json

This is a work in progress.

Many of the endpoints work, but not all features have been implemented, in particular filter terms for computation queries, and more testing is needed.

At present, all metadata are saved in the pre-production version of the KG, which is reset from time-to-time, and so metadata will not be preserved long-term: for now, please use this only for testing.

Authorize 

### Authentication and authorization

**GET** / About This Api 

**GET** /login Login Via Ebrains 

**GET** /auth Auth Via Ebrains 

### Simulations

**GET** /simulations/ Query Simulations  

**POST** /simulations/ Create Simulation  

**GET** /simulations/{simulation\_id} Get Simulation  

**PUT** /simulations/{simulation\_id} Replace Simulation  

**DELETE** /simulations/{simulation\_id} Delete Simulation  

<https://prov.brainsimulation.eu/docs>

# Using the Provenance API

- “manual” use by individual users, e.g. in Jupyter notebooks
- integration into apps & services, e.g.
  - track simulations in Model Catalog app
  - migrate Neuromorphic job queue metadata to KG
  - capture all experiments run on the Neurorobotics Platform?
  - integrate into HPC job proxy?
- UNICORE integration / wrapper
- integration into CWL runners
- workflows dashboard
- provenance tracking of the Spack / Docker build process?!

# Demo

---

<https://wiki.ebrains.eu/bin/view/Collabs/ebrains-workflows>

.../examples/visualisation/plot\_ephys\_data\_with\_prov\_api.ipynb

# Using / visualising provenance information

**EBRAINS: Data analysis and simulation pipelines**

Object class  Object ID  **SEARCH**

<https://prov-vis.brainsimulation.eu/>

**MultiChannelMultiTrialRecording**  
Traces recorded in WBS\_WT\_S5\_samp-7 Screenshot  
<https://nexus.humanbrainproject.org/v0/data/neuralactivity/electrophysiology/multitrace/v0.2.0/542e63ea-e09e-415b-9ca7-36e37c4fe332>

AVAILABLE PIPELINES

- Multi-channel ECoG ...  
02/06/2020, 08:17:41
- Multi-channel ECoG ...  
31/05/2020, 15:52:54
- Multi-channel ECoG ...**  
31/05/2020, 10:55:57
- Multi-channel ECoG ...  
29/05/2020, 20:02:43
- Multi-channel ECoG ...  
29/05/2020, 15:43:46
- Multi-channel ECoG ...  
29/05/2020, 13:45:24
- Multi-channel ECoG ...  
29/05/2020, 13:32:11
- Multi-channel ECoG ...

**ANALYSISRESULT**  
**Multi-channel ECoG ...**  
31/05/2020, 10:55:57  
Andrew Davison  
scripts/curate\_IDIBAPS\_data.py  
NIX

**ANALYSISRESULT**  
**Figure, generated f...**  
31/05/2020, 10:56:39  
Andrew Davison  
scripts/plot\_traces.py  
PNG

**ANALYSISRESULT**  
**Frame array, genera...**  
31/05/2020, 11:02:02  
Andrew Davison  
scripts/background\_substraction.py  
NPY

**ANALYSISRESULT**  
**Figure, generated f...**  
31/05/2020, 11:02:02  
Andrew Davison  
scripts/background\_substraction.py  
PNG

**ANALYSISRESULT**  
**Multi-channel ECoG ...**  
31/05/2020, 11:02:03  
Andrew Davison  
scripts/background\_substraction.py  
NIX